Statistica Sinica **27** (2017), 1385-1400
doi:https://doi.org/10.5705/ss.202015.0219

SIMULTANEOUS CONFIDENCE BANDS IN NONLINEAR REGRESSION MODELS WITH NONSTATIONARITY

Degui Li, Weidong Liu, Qiying Wang and Wei Biao Wu

The University of York, Shanghai Jiao Tong University, The University of Sydney and The University of Chicago

Abstract: We consider nonparametric estimation of the regression function $g(\cdot)$ in a nonlinear regression model $Y_t = g(X_t) + \sigma(X_t)e_t$, where the regressor (X_t) is a nonstationary unit root process and the error (e_t) is a sequence of independent and identically distributed (i.i.d.) random variables. With proper centering and scaling, the maximum deviation of the local linear estimator of the regression function g is shown to be asymptotically Gumbel. Based on the latter result, we construct simultaneous confidence bands for g, which can be used to test patterns of the regression function. Our results extend existing ones that typically require independent or stationary weakly dependent regressors. We examine the finite sample behavior of the proposed approach via simulated and empirical data examples.

Key words and phrases: Gumbel convergence, integrated process, local linear estimation, local time limit theory, maximum deviation, simultaneous confidence bands.

1. Introduction

Following Granger (1981) and Engle and Granger (1987), cointegration has become one of the most important topics in econometrics. For non-stationary processes (X_t) and (Y_t), we say that they are (linearly) cointegrated if there exists a nonzero constant a such that

$$Y_t - aX_t = \varepsilon_t, \quad t = 1, \ldots, n, \tag{1.1}$$

where (ε_t) is a stationary process. In essence, the nonstationary processes (X_t) and (Y_t) have a common stochastic trend over a long time series period. The classic linear cointegration models have been extensively studied in the literature. To account for possible nonlinear relations, we consider the nonlinear and nonstationary regression model

$$Y_t - g(X_t) = \sigma(X_t)e_t, \quad t = 1, \ldots, n, \tag{1.2}$$

where (e_t) is a stationary error process, (X_t) is a nonstationary regressor, g and σ are smooth functions. Model (1.2) generalizes (1.1) by allowing possible nonlinear structures on both g and σ. It provides a flexible tool to analyze nonlinear relationships between two nonstationary processes.

The aim of this paper is to study nonparametric estimates of the function g in model (1.2). Convergence properties of the conventional Nadaraya-Watson kernel-based estimator have been widely discussed under the assumption that (X_t, Y_t) are i.i.d.; see Härdle (1990) and the references therein. In Györfi et al. (1989); Bosq (1996) and Fan and Yao (2003), the i.i.d. assumption is relaxed and strong mixing stationary processes are allowed. The asymptotic problem is quite challenging if (X_t) is non-stationary. Karlsen, Myklebust and Tjøstheim (2007, 2010) assumed that (X_t) is a null recurrent Markov chain. Wang and Phillips (2009a,b, 2011) and Cai, Li and Park (2009) considered the integrated process $X_t = \sum_{j=1}^{t} x_j$, where (x_j) is a stationary linear process. For other literature on this research area, we refer to Gao et al. (2009a,b); Chen, Li and Zhang (2010); Chen, Gao and Li (2012); Kasparis and Phillips (2012); Wang and Phillips (2012); Wang (2014).

The aforementioned papers deal with the point-wise central limit theorem for the Nadaraya-Watson estimator. This type of result, however, is not useful for testing certain patterns of the regression function g. In recent years, there has been increasing interest in deriving uniform consistency results for the kernel-based estimator under a nonstationary framework. For instance, Wang and Wang (2013) and Chan and Wang (2014) established uniform consistency of the kernel-based estimator by using local time limit theory, Wang and Chan (2014) obtained uniform convergence rates for a class of martingales and studied their application in nonlinear cointegration regression, and Gao et al. (2015) derived some uniform consistency results using the framework of null recurrent Markov chains.

In this paper, we estimate the regression function g using the local linear smoothing method in which the estimator of $g(x)$ is

$$\widehat{g}_n(x) = \frac{\sum_{t=1}^{n} w_t(x) Y_t}{\sum_{t=1}^{n} w_t(x)}, \tag{1.3}$$

where the weight function is

$$w_t(x) = K[\frac{(X_t - x)}{h}]V_{n2}(x) - K_1[\frac{(X_t - x)}{h}]V_{n1}(x)$$

with $V_{nj}(x) = \sum_{i=1}^{n} K_j[(X_i - x)/h]$, $K_j(x) = x^j K(x)$, $K(\cdot)$ being a non-negative real function, and the bandwidth $h \equiv h_n \to 0$. Under the stationarity assumption on the observations, existing literature such as Fan and Gijbels (1996) has shown that local linear estimation has some advantages over the Nadaraya-Watson kernel estimation method. A recent paper by Chan and Wang (2014) further showed that the performance of $\widehat{g}_n(\cdot)$ is superior to that of the conventional Nadaraya-Watson estimator in uniform asymptotics for nontationary time series.

To assess patterns of the regression function g, for example, to test whether g is linear or of other parametric forms, we need to construct a simultaneous confidence band (SCB) for g over a suitable interval. Neither the point-wise central limit theorem nor the uniform convergence of $\widehat{g}_n(\cdot)$ is sufficient for testing whether g has a particular functional form. Here we obtain the asymptotic distribution of the normalized maximum absolute deviation

$$\Delta_n = \sup_{|x|\le B_n} \left| V_n(x) \frac{\left[\widehat{g}_n(x) - g(x)\right]}{\sigma(x)} \right|, \tag{1.4}$$

where B_n is a sequence of positive constants which may diverge to infinity, and the normalizing term is

$$V_n(x) = V_{n2}^{-1}(x) \frac{\sum_{t=1}^n w_t(x)}{\left(\sum_{t=1}^n K^2\left[(X_t - x)/h\right]\right)^{1/2}}. \tag{1.5}$$

Such an asymptotic distributional theory refines the existing uniform consistency results such as those obtained by Chan and Wang (2014), and it further enables one to construct a SCB for the unknown regression function g. In the traditional simultaneous inference theory it is assumed that the regressor process (X_t) is i.i.d. or stationary; see, for example, Bickel and Rosenblatt (1973); Johnston (1982); Hall and Titterington (1988); Fan and Zhang (2000); Zhao and Wu (2008); Liu and Wu (2010); Xia (1998) and Zhang and Peng (2010). However, in our setting, due to the nonstationarity and the dependence, it is very demanding to establish a limit theory for Δ_n. To this end, we introduce new technical mechanisms and obtain a precise characterization of $V_n(x)$ over an unbounded interval.

The rest of the paper is organized as follows. The assumptions and main theoretical results are stated in Section 2. By using the asymptotic distribution of Δ_n, in Section 3 we construct SCBs for the regression function g over an expanding interval. In Section 4, we provide simulated and empirical data examples to illustrate the finite sample behavior of the proposed approach. The proofs of the main results are in Section 5. Some technical lemmas with proofs, and some supplemental asymptotic theorems are given in a supplemental document.

2. Main Results

We need some regularity conditions to establish the asymptotic distribution of Δ_n as (1.4). Let C denote a generic positive constant.

(C1) [Regressor process] Let

$$X_t = \sum_{j=1}^{t} x_j, \quad x_j = \sum_{k=0}^{\infty} \phi_k \, \eta_{j-k},$$

where (η_j) is a sequence of i.i.d. random variables with $\mathsf{E}[\eta_1] = 0$, $\mathsf{E}[\eta_1^2] = 1$, and $\mathsf{E}[|\eta_1|^{2+\delta}] < \infty$ for some $\delta > 0$; the characteristic function $\varphi(t)$ of η_1 satisfies $\int_{-\infty}^{\infty}(1+|t|)|\varphi(t)|dt < \infty$, and the coefficients $(\phi_k)_{k\geq 0}$ satisfy $\sum_{k=0}^{\infty}|\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty}\phi_k \neq 0$.

(C2) [Regression function] The first derivative $g'(x)$ of $g(x)$ exists. There is a real positive function $g_0(x)$ such that

$$|g'(y) - g'(x)| \leq g_0(x)|y - x|, \tag{2.1}$$

uniformly for $x \in \mathsf{R}$ and $|y - x|$ sufficiently small.

(C3) [Range and Bandwidth] (i) $B_n = M_0 n^{1/2-\epsilon}$, where M_0 is a constant and $\epsilon > 0$ can be arbitrarily small; (ii) for some constant $0 < \delta_0 < 1/4$, $n^{1/2-\delta_0}h \to \infty$; (iii) $nh^{10}\sup_{|x|\leq 2B_n}\left[1 + g_0^4(x)\right] = O(\log^{-8} n)$, where $g_0(x)$ is given in (2.1).

(C4) [Errors] (i) The error process (e_t) is i.i.d. with $\mathsf{E}[e_1] = 0$, $\mathsf{E}[e_1^2] = 1$ and $\mathsf{E}[|e_1|^{2p}] < \infty$, where $p \geq 1 + [1/\delta_0]$ and δ_0 is defined as in (**C3**)(ii), and (e_t) is independent of the process (η_t). (ii) For the function $\sigma(\cdot)$,

$$\inf_{x\in\mathsf{R}} \sigma(x) > 0, \quad \sup_{x\in\mathsf{R}} \frac{|\sigma(x+y) - \sigma(x)|}{\sigma(x)} \leq C|y| \tag{2.2}$$

for any $|y|$ sufficiently small.

(C5) [Kernel] The kernel function K is absolutely continuous on a compact support $[-A, A]$ with $A > 0$, $\int xK(x)dx = 0$, and $|K(x) - K(y)| \leq C|x - y|$ for all $x, y \in \mathsf{R}$.

Remark 1. Conditions similar to **(C1)** and **(C2)** are common; see, for instance, Wang and Phillips (2009b); Wang and Wang (2013), and Chan and Wang (2014). The smoothness condition (2.1) on g' is slightly weaker than the existence of second derivative on g. If the function g has continuous and bounded second order derivative, we may replace $g_0(x)$ in (2.1) by a positive constant and **(C3)** (iii) can be simplified to $nh^{10}\log^8 n = O(1)$.

In condition **(C3)**(i), we allow B_n to diverge to infinity, which is an essential difference from the literature that investigates the SCB for stationary regressor with B_n usually assumed to be fixed (c.f., Liu and Wu (2010)). Our main theorem is established for diverging B_n, which indicates that the SCB for unknown g with $I(1)$ regressor may be constructed under an expanding interval whose length is allowed to diverge to infinity; see Remark 4 in Section 3 for more details. **(C3)**(ii) is close to the necessary condition of $\sqrt{n}h \to \infty$ by letting δ_0 be sufficiently close to zero. **(C3)**(iii) implies that there is a trade-off between the function g, the bandwidth h, and the range B_n. If B_n is assumed to be a positive constant B, **(C3)**(iii) is satisfied when $\sup_{|x|\le 2B} g_0(x) = O(1)$ and $nh^{10}\log^8 n = O(1)$. In particular, the condition **(C3)**(iii) ensures that the bias term of the local linear estimator (1.3) with nonstationarity is asymptotically negligible.

In condition **(C4)**, $\mathsf{E}[e_1^2] = 1$ is not necessary. If $\mathsf{E}[e_1^2] = a^2 \neq 1$, it suffices to standardize the model (1.2) with $e_t' = e_t/a$ and replace $\sigma(x)$ by $a\sigma(x)$. The moment condition of e_1 depends on the δ_0 in **(C3)**(ii), which is reasonable. Weaker bandwidth restriction (smaller δ_0) would lead to stronger moment condition on e_t (larger p). We require the independence between the regressor (X_t) and the error (e_t) by noting that (e_t) is independent of (η_t) in **(C4)**(i). This condition is restrictive, but seems difficult to be relaxed, even with a stationary regressor. The condition imposed on $\sigma(x)$ is mild and is satisfied by a large class of functions. Typical examples include $\sigma(x) = 1 + |x|^k$ with $k \ge 0$ and $\sigma(x) = e^x/(1+e^x)$.

The kernel condition **(C5)** implies that the derivative $K'(x)$ exists almost everywhere and $\int_{-A}^{A}[K'(x)]^2dx < \infty$. Let $\lambda_1 = \int_{-A}^{A} K(x)dx$, $\lambda_2 = \int_{-A}^{A} K^2(x)dx$, and $\kappa_2 = \int_{-A}^{A}(K'(x))^2dx/(2\lambda_2)$. We further have

$$r(s) \equiv \int_{-A}^{A} K(x)K(x+s)\frac{dx}{\lambda_2} = 1 - \kappa_2 s^2 + o(s^2), \qquad \text{as } s \to 0. \tag{2.3}$$

This follows from the Taylor's expansion of $r(s)$ and Theorems B1–B2 of Bickel and Rosenblatt (1973). As there, $r(s)$ is used as the covariance function of a certain Gaussian process. The kernel condition **(C5)** is mild and it is satisfied by such commonly-used functions as the Epanechnikov kernel $K(u) = \max\{3(1-u^2)/4, 0\}$.

Theorem 1. *Let Conditions* (C1)-(C5) *be satisfied. Then, for* $z \in \mathsf{R}$,

$$\mathsf{P}\left\{(2\log \bar{h}^{-1})^{1/2}\,(\Delta_n - d_n) \le z\right\} \to e^{-2e^{-z}}, \tag{2.4}$$

where $\bar{h} = h/(2B_n)$ *and*

$$d_n = (2\log \bar{h}^{-1})^{1/2} + \frac{1}{(2\log \bar{h}^{-1})^{1/2}} \log \frac{\kappa_2^{1/2}}{\pi 2^{1/2}}.$$

Remark 2. If $\mathcal{G}_n = (2\log \bar{h}^{-1})^{1/2}\,(\Delta_n - d_n)$ and $\mathcal{G}$ is the standard Gumbel distribution with the cumulative distribution function defined as the limit in (2.4), we can reformulate the above limit result as $\mathcal{G}_n \stackrel{d}{\to} \mathcal{G}$. As we allow that B_n is divergent, using Theorem 1 we can construct SCBs for the unknown function g on expanding intervals; the requirement of intervals being expanding is crucial in our functional estimation since the regressor process (X_t) is integrated and thus is stochastically unbounded, behaving like a random walk. Liu and Wu (2010) obtained Gumbel convergence for the nonlinear regression models with stationary regressors, but assumed that the involved interval is bounded and non-expanding.

The asymptotic Gumbel distribution result (2.4) also holds when we replace $\widehat{g}_n(\cdot)$ by the Nadaraya-Watson kernel estimation $\widetilde{g}_n(\cdot)$ in the definition of Δ_n. We provide some relevant results in Appendix C of the supplemental document.

3. Construction of SCBs

This section constructs the SCBs for the regression function g. Since $\sigma(x)$ in the definition of Δ_n given in (1.4) is unknown, a consistent estimate of $\sigma(x)$ satisfying certain rates is required over the set $\{x : |x| \le B_n\}$. Using arguments as in Wang and Wang (2013), we construct the kernel estimate

$$\widehat{\sigma}_n^2(x) = \frac{\sum_{t=1}^n [Y_t - \widehat{g}_n(X_t)]^2 K[(X_t - x)/b]}{\sum_{t=1}^n K[(X_t - x)/b]}, \tag{3.1}$$

where $\widehat{g}_n(x)$ is the local linear estimate defined in (1.3), and b is a bandwidth. Let $a_n \asymp b_n$ signify that a_n and b_n have the same asymptotic order.

Proposition 1. *Under the conditions of Theorem 1 and $b \asymp h$, we have*

$$\sup_{|x|\le B_n} \left| \frac{\sigma(x)}{\widehat{\sigma}_n(x)} - 1 \right| = O_P\big[h + (nh^2)^{-1/4} \log^2 n\big]. \tag{3.2}$$

Remark 3. The proof of Proposition 1 is given in Section 5. Proposition 1 can be seen as an extension of Wang and Wang (2013)'s uniform consistency results from the case of bounded range to the case of diverging range. The uniform convergence rate in (3.2) is close to optimal. Using the bandwidth conditions in **(C3)**, the rate in (3.2) is sufficient for replacing $\sigma(\cdot)$ by $\widehat{\sigma}_n(\cdot)$ when constructing the SCBs of the regression function g.

Let

$$\widehat{\Delta}_n = \sup_{|x|\le B_n} \left| V_n(x) \, \frac{\big[\widehat{g}_n(x) - g(x)\big]}{\widehat{\sigma}_n(x)} \right|.$$

Due to (3.2) and Theorem 1, $\widehat{\Delta}_n$ and Δ_n have the same limit distribution. Consequently, for given α, the $(1-\alpha)$-SCB for g over the set $\{x : |x| \le B_n\}$ can be constructed as

$$\left[\widehat{g}_n(x) - l_\alpha(x), \quad \widehat{g}_n(x) + l_\alpha(x)\right], \tag{3.3}$$

where

$$l_\alpha(x) = \left[z_\alpha (2\log \bar{h}^{-1})^{-1/2} + d_n\right] \widehat{\sigma}_n(x)\, V_n^{-1}(x),$$
$$z_\alpha = -\log\left(-\frac{1}{2}\log(1-\alpha)\right), \quad \bar{h} = \frac{h}{(2B_n)},$$
$$d_n = (2\log \bar{h}^{-1})^{1/2} + \frac{1}{(2\log \bar{h}^{-1})^{1/2}} \log \frac{\kappa_2^{1/2}}{\sqrt{2}\pi}.$$

Remark 4. As Theorem 1 implies that the asymptotic bias term is negligible due to the bandwidth condition **(C3)**, we do not need to correct the bias of the local linear estimation when constructing the SCBs in (3.3). The simulation study in Section 4 shows that such construction of SCBs works reasonably well in the finite sample case. When the regression function $g(x)$ has a thin tail, such as $g(x) = (\alpha + \beta e^x)/(1+e^x)$, some routine calculations show that $\sup_x |g_0(x)| < \infty$. As a consequence, the SCB result (3.3) holds over the set $\{x : |x| \le M_0 n^{1/2-\epsilon}\}$ whenever $n^{1/2-\delta_0} h \to \infty$ and $nh^{10}\log^8 n \to 0$, where ϵ can be chosen sufficiently small.

4. Numerical Studies

In this section, we report on simulated and empirical data examples to illustrate the finite sample behavior of our SCBs. For the construction, we need to choose an appropriate cut-off value so that the pre-assigned nominal confidence level can be achieved. The Gumbel convergence in Theorem 1 can be quite slow, which implies that the SCB in (3.3) using the asymptotical cut-off value may not have good finite sample performance. To circumvent such a problem, as in Wu and Zhao (2007) and Liu and Wu (2010), we introduce a simulation-based method.

1. Choose an appropriate bandwidth h and a kernel function K satisfying **(C5)**. Then using (1.3), compute $\widehat{g}_n(x)$ and the estimated residuals $\widehat{e}_t = Y_t - \widehat{g}_n(X_t)$. Based on the latter, with another bandwidth b, estimate $\sigma^2(\cdot)$ by using (3.1) and denote the estimate by $\widehat{\sigma}_n^2(\cdot)$.
2. Generate i.i.d. standard normal random variables $e_1^*, \ldots, e_n^*$ independent of $x_1, \ldots, x_n$; compute $Y_t^* = \widehat{\sigma}_n(X_t) e_t^*$; perform a local linear regression of (Y_t^*) on (X_t) with bandwidth h and kernel K, and let $\hat{g}_n^*(\cdot)$ be the estimated

regression function. Then compute the normalized maximum deviation

$$\Delta_n^* = \max_{|x|\le B_n} \Big|\frac{V_n(x)\hat{g}_n^*(x)}{\widehat{\sigma}_n(x)}\Big|. \tag{4.1}$$

3. Repeat Step 2 for N times, and then compute the $(1-\alpha)$th sample quantile, denoted by $\Delta_{n,1-\alpha}^*$.

4. Construct the $(1-\alpha)$th SCB for $g(x)$ over the interval $x \in [-B_n, B_n]$ as

$$\left[\widehat{g}_n(x) - \widehat{\sigma}_n(x)\frac{\Delta_{n,1-\alpha}^*}{V_n(x)},\ \ \widehat{g}_n(x) + \widehat{\sigma}_n(x)\frac{\Delta_{n,1-\alpha}^*}{V_n(x)}\right].$$

From Theorem 1, with the normalization in (2.4), Δ_n^* and Δ_n have the same asymptotic Gumbel distribution. Then the empirical quantile of the former can approximate that of the latter. In comparison with the asymptotic Gumbel distribution, the distribution of Δ_n^* better approximates that of Δ_n. Our simulation-based construction has an important practical convenience: the procedure is the same as when data are stationary. We do not have to be concerned with whether the true data generating process is $I(0)$ or $I(1)$ for constructing SCBs with asymptotically correct coverage probabilities.

Example 1. We took the nonlinear regression model

$$Y_t = g(X_t) + \sigma(X_t)e_t, \quad X_t = X_{t-1} + x_t, \quad t = 1, 2, \cdots, n, \tag{4.2}$$

with $g(x) = \log(10 + 0.25x^{1.2})$, $\sigma(x) = (6 + 0.2x^2)^{0.5}$, (x_t) generated by the AR(1) process $x_t = 0.4x_{t-1} + v_t$, $v_t, t \in \mathsf{Z}$, i.i.d. $\mathsf{N}(0,1)$, e_t, $t \in \mathsf{Z}$, i.i.d. uniform$(-3^{1/2}, 3^{1/2})$, and $(v_t)_{t\in\mathsf{Z}}$ and $(e_t)_{t\in\mathsf{Z}}$ independent. We chose $n = 100$, 200, 500 and considered SCBs over the interval $x \in [B_l, B_u]$, where B_l (resp. B_u) is the 0.1th (resp. 0.9th) sample quantile of the $I(1)$ regressor (X_t).

In our simulation, we used the Gaussian kernel function $K(u) = (2\pi)^{-1/2}$ $\exp(-u^2/2)$, since it has a very thin tail and is the default choice in R package KernSmooth. We used the function dpill (Ruppert, Sheather and Wand (1995)) in the package KernSmooth for choosing the bandwidth and the function locpoly for performing the local linear regression. Table 1 shows the coverage probabilities as proportions of the SCBs that cover the true function $g(x)$ over $x \in [B_l, B_u]$ based on 10^5 repetitions. The levels considered were: $\alpha = 0.05$ and $\alpha = 0.01$. Coverage probabilities based on theoretical cut-off values computed from Theorem 1 are also shown. In the left panel we use the true variance function $\sigma^2(\cdot)$, while in the right one the estimated $\widehat{\sigma}_n^2(\cdot)$ is used. Using the true variance function $\sigma^2(\cdot)$, with $\alpha = 5\%$ and $n = 500$, the coverage probability for the simulation-based method is 0.9495. For the theoretical cut-off value based on Theorem 1, the coverage

Table 1. Coverage probabilities for SCBs.

	True $\sigma^2(\cdot)$		Estimated $\sigma^2(\cdot)$	
$(\alpha; n)$	Simulation	Theoretical	Simulation	Theoretical
0.05; 100	0.9411	0.9568	0.8859	0.9069
0.01; 100	0.9874	0.9971	0.9564	0.9805
0.05; 200	0.9376	0.9530	0.9098	0.9271
0.01; 200	0.9875	0.9960	0.9723	0.9878
0.05; 500	0.9495	0.9622	0.9447	0.9573
0.01; 500	0.9901	0.9974	0.9880	0.9963

probability is 0.9622. As expected, larger n leads to more accurate coverage probabilities. A similar claim can be made for the SCBs with $\alpha = 1\%$. Hence the simulation-based method has a more accurate finite sample performance, in particular when $n = 500$. The accuracy of the coverage probabilities can be slightly affected if the estimated variance function $\widehat{\sigma}_n^2(\cdot)$ is used.

Example 2. We considered the monthly US share price indices and treasury bill rates for the period January/1957–December/2009, downloaded from International Monetary Fund's website, as in Chen, Gao and Li (2012). The upper two plots in Figure 1 give the two series with the log transformation applied to the share price data. The augmented Dickey-Fuller test indicates that the treasury bill rates are the $I(1)$ process, and the log-transformed share price indices are the $I(1)$ process with drift. To ensure that our methodology and theory are applicable to the data, we removed the drift from the share price indices and the adjusted share price indices are plotted in the lower-left plot of Figure 1. The lower-right plot in Figure 1 gives the scatter plot of the adjusted share price indices against the treasury bill rates, which indicates the existence of heteroskedasticity. The aim of this example is to analyze the relationship between the share price series and the treasury bill rates. Let (Y_t) be the adjusted share price series, and (X_t) be the treasury bill series. We first fit the data with the linear regression model $Y_t = a + bX_t + e_t$, and obtained the least squares estimation $\widehat{a}$ and $\widehat{b}$ of the two parameters: -0.1582 and -0.0514, respectively. The cointegration test shows that the null hypothesis of the $I(1)$ error process cannot be rejected. This indicates that the linear cointegration model is not appropriate for the data.

We considered the nonlinear regression model

$$Y_t = g(X_t) + \varepsilon_t^*, \quad \varepsilon_t^* = \sigma(X_t)e_t, \quad t = 1, \cdots, 636, \tag{4.3}$$

where (e_t) is assumed to be independent of (X_t), g and σ are two nonlinear

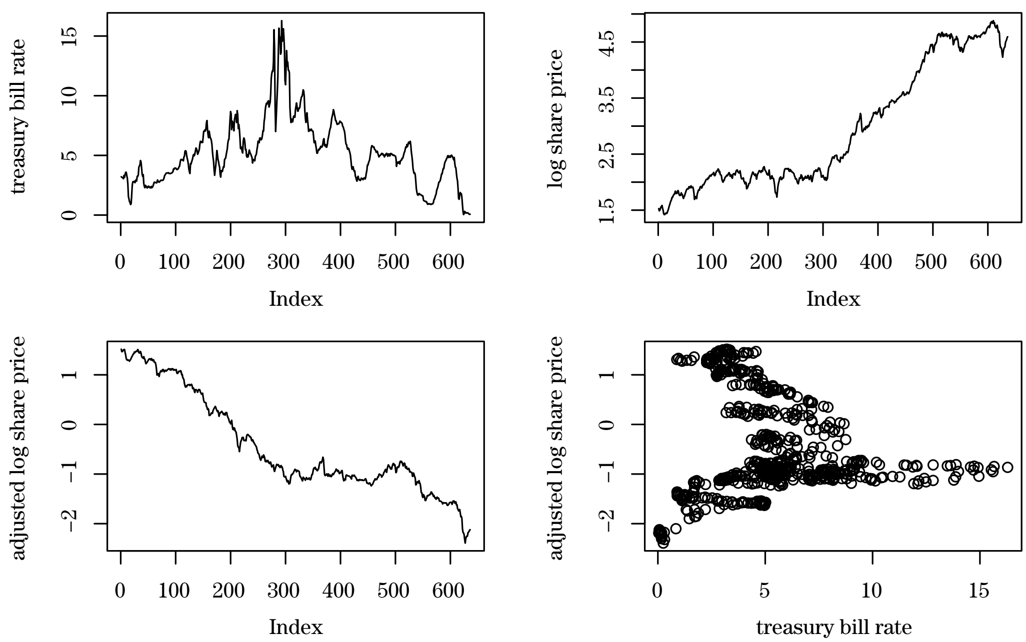

Figure 1. The data plots over period of January/1959–December/2009.

functions as in model (1.2). The structure of $\varepsilon_t^* = \sigma(X_t)e_t$ allows the existence of nonstationarity and heteroskedasticity for the model error term. The local linear estimation method was used to estimate the regression function g, with resulting estimate $\widehat{g}_n(\cdot)$, where the bandwidth was chosen by the commonly-used cross-validation method. Let $\widehat{\varepsilon}_t^* = Y_t - \widehat{g}_n(X_t)$. Figure 2 plots the residual $(\widehat{\varepsilon}_t^*)$ against the regressor (X_t); it shows that our nonlinear model (4.3) with heteroskedasticity on errors is appropriate for the data.

The SCB for the regression function is plotted in Figure 3 with the cut-off value chosen by the simulation-based method. The upper and lower dashed lines are the 95%-SCB for the regression function g. The solid line is the local linear estimated regression function and the dotted line is the estimated linear regression function. From the figure, we have to reject the hypothesis that g is linear at 5% level as part of the dotted line lies outside the SCB. This again shows that (X_t) and (Y_t) are not linearly cointegrated and the traditional linear model is not suitable for the data.

It is difficult to derive the local time limit theory for local linear estimation when (X_t) is multivariate $I(1)$, which limits the empirical applicability of the proposed SCB construction. While Figures 3 suggests a nonlinear relationship between the adjusted share price series and the treasury bill series, the true

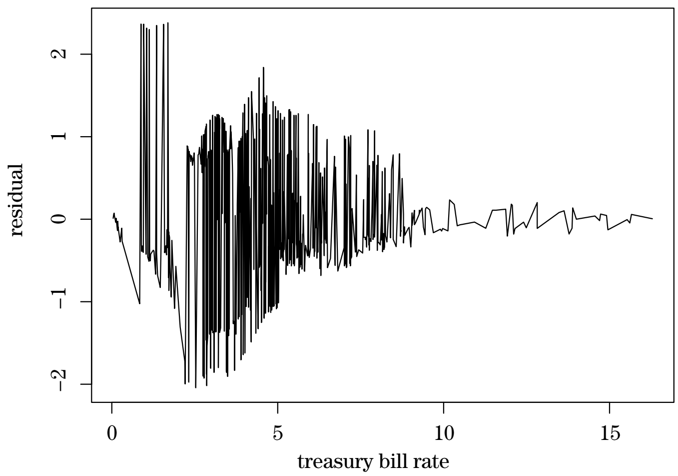

Figure 2. The residuals from the nonlinear regression model against the treasury bill rates.

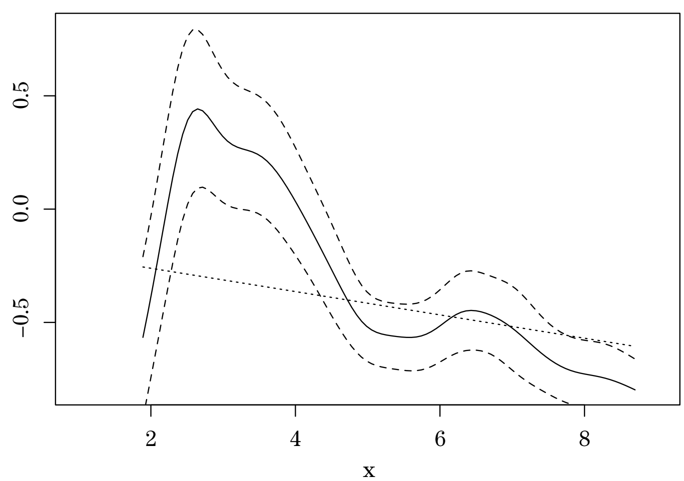

Figure 3. The SCB for the regression function g.

relationship may also depend on some other macroeconomic variables such as the long-term bond yields (Chen, Gao and Li (2012)). To address the latter problem, we could extend the nonlinear regression model (4.3) to the nonlinear varying-coefficient models (c.f., Cai, Li and Park (2009)) and include some other macroeconomic time series in the empirical analysis. This is beyond the scope of this paper.

5. Proofs of the Main Results

In this section, we give the proofs of Theorem 1 in Section 2 and Proposition 1 in Section 3. We prove Theorem 1 for $B_n = \sqrt{n}/(\log^{c_0} n)$ with c_0 being a positive constant, which is asymptotically wider than $B_n = M_0 n^{1/2-\epsilon}$. Some of the arguments used in this section and the supplemental material are similar to those in Bickel and Rosenblatt (1973) and Liu and Wu (2010). However, due to the presence of non-stationary series, our proofs are quite challenging.

Proof of Theorem 1. Let

$$\Gamma_{1n}(x) = \frac{1}{S_n(x)V_{n2}(x)} \cdot \sum_{t=1}^{n} w_t(x)\big[g(X_t) - g(x)\big],$$

$$\Gamma_{2n}(x) = \frac{V_{n1}(x)}{S_n(x)V_{n2}(x)} \cdot \sum_{t=1}^{n} K_1[\frac{(X_t - x)}{h}]e_t,$$

$$\Gamma_{3n}(x) = \frac{1}{S_n(x)V_{n2}(x)} \cdot \sum_{t=1}^{n} w_t(x)\big[\sigma(X_t) - \sigma(x)\big]\frac{e_t}{\sigma(x)},$$

where $S_n^2(x) = \sum_{t=1}^n K^2\big[(X_t - x)/h\big]$, $V_{nj}(\cdot)$, $w_t(\cdot)$ and $K_1(\cdot)$ are defined as in Section 1. Similarly to Liu and Wu (2010), we split the $V_n(x)\big[\widehat{g}_n(x) - g(x)\big]/\sigma(x)$ as

$$\begin{aligned} V_n(x)\frac{\big[\widehat{g}_n(x) - g(x)\big]}{\sigma(x)} &= V_n(x)\frac{\Big\{\sum_{t=1}^n w_t(x)\sigma(X_t)e_t + \sum_{t=1}^n w_t(x)\big[g(X_t) - g(x)\big]\Big\}}{\big[\sigma(x)\sum_{t=1}^n w_t(x)\big]} \\ &= \frac{1}{S_n(x)} \cdot \sum_{t=1}^{n} K[\frac{(X_t - x)}{h}]e_t + \frac{\Gamma_{1n}(x)}{\sigma(x)} + \Gamma_{2n}(x) + \Gamma_{3n}(x). \end{aligned} \tag{5.1}$$

We apply truncation to deal with the first term on the right hand side of (5.1). Let $\mathcal{I}_n = [-h^{-1}B_n, h^{-1}B_n]$ and $e_t' = \tilde{e}_t/(\mathsf{E}\tilde{e}_t^2)^{1/2}$, where

$$\tilde{e}_t = e_t I\{|e_t| \le \log n\} - \mathsf{E}\left[e_t I\{|e_t| \le \log n\}\right], \quad 1 \le t \le n. \tag{5.2}$$

Take $Z_t(x) = K(X_t/h - x)/S_n(xh)$ and

$$M_n(x) = \sum_{t=1}^{n} Z_t(x)e_t, \quad M_n = \sup_{x\in\mathcal{I}_n} |M_n(x)|;$$

$$\widetilde{M}_n(x) = \sum_{t=1}^{n} Z_t(x)e_t', \quad \widetilde{M}_n = \sup_{x\in\mathcal{I}_n} |\widetilde{M}_n(x)|.$$

The main idea is to show that $\Gamma_{jn}(x)$, $j = 1, 2, 3$, are asymptotically dominated by $M_n(x)$ uniformly over $|x| \leq B_n$, and $M_n(x)$ is asymptotically equivalent to its truncated version $\widetilde{M}_n(x)$. Due to (5.1) and $\inf_{x\in \mathsf{R}} \sigma(x) > 0$, Theorem 1 follows from the three propositions. Their proofs are in the supplemental materials.

Proposition 2. *Under the conditions of Theorem 1, we have*

$$\sup_{|x|\leq B_n} |\Gamma_{jn}(x)| = O_P(\log^{-2} n), \quad j = 1, 2, 3.$$

Proposition 3. *Under the conditions of Theorem 1, we have*

$$\sup_{x\in\mathcal{I}_n} |M_n(x) - \widetilde{M}_n(x)| = O_P(\log^{-2} n).$$

Proposition 4. *Under the conditions of Theorem 1, we have for any $z \in R$,*

$$\mathsf{P}\Big\{(2\log \bar{h}^{-1})^{1/2}(\widetilde{M}_n - d_n) \leq z\Big\} \to e^{-2e^{-z}},$$

where d_n is defined as in Theorem 1.

Proof of Proposition 1. Due to the condition (C4)(ii) and $K(s) = 0$ if $|s| \geq A$,

$$\frac{\big|\sigma^i(X_t) - \sigma^i(x)\big|}{\sigma^i(x)} \cdot K\big[\frac{(X_t - x)}{b}\big] \leq CbK\big[\frac{(X_t - x)}{b}\big],$$

for $i = 1$ and 2, all $x \in \mathsf{R}$, and a sufficiently small b. Similarly, we have

$$|\widehat{g}_n(X_k) - g(X_k)|^i\, K\big[\frac{(X_k - x)}{b}\big] \leq CK\big[\frac{(X_k - x)}{b}\big] \cdot \sup_{|x|\leq 2B_n} |\widehat{g}_n(x) - g(x)|^i\,,$$

for $i = 1$ and 2, $|x| \leq B_n$, and a sufficiently small b. Then, we have

$$\begin{aligned}\frac{\big|\widehat{\sigma}_n^2(x) - \sigma^2(x)\big|}{\sigma^2(x)} \leq\; & C\left[b + \sigma^{-2}(x) \sup_{|z|\leq 2B_n} |\widehat{g}_n(z) - g(z)|^2\right] + \\ & C\left|\frac{\sum_{t=1}^n K\big[(X_t - x)/b\big](e_t^2 - 1)}{\sum_{t=1}^n K\big[(X_t - x)/b\big]}\right| + \\ & C\left[\frac{\sum_{t=1}^n K\big[(X_t - x)/b\big]|e_t|}{\sigma(x)\sum_{t=1}^n K\big[(X_t - x)/b\big]} \cdot \sup_{|z|\leq 2B_n} |\widehat{g}_n(z) - g(z)|\right],\end{aligned}$$

for $|x| \leq B_n$ and a sufficiently small b. By Theorem 4.1 of Chan and Wang (2014) and (C3)(iii), we have

$$\begin{aligned}\sup_{|x|\leq 2B_n} |\widehat{g}_n(x) - g(x)| &= O_P\big[(nh^2)^{-1/4}\log^{1/2} n + h^2 \sup_{|x|\leq 2B_n} |g_0(x)|\big] \\ &= O_P\big[(nh^2)^{-1/4}\log^2 n\big].\end{aligned}$$

The above arguments, together with Lemma A.4 given in the supplemental document, $b \asymp h$, and $\inf_{x\in\mathsf{R}} \sigma(x) > 0$, lead to

$$\sup_{|x|\le B_n} \frac{\left|\widehat{\sigma}_n^2(x) - \sigma^2(x)\right|}{\sigma^2(x)} = O_P\left[h + (nh^2)^{-1/4}\log^2 n\right].$$

Hence, we have

$$\sup_{|x|\le B_n} \left|\frac{\sigma(x)}{\widehat{\sigma}(x)} - 1\right| \le \sup_{|x|\le B_n} \frac{\sigma(x)}{\widehat{\sigma}(x)} \frac{\left|\widehat{\sigma}^2(x) - \sigma^2(x)\right|}{\sigma^2(x)} = O_P\left[h + (nh^2)^{-1/4}\log^2 n\right].$$

This completes the proof of Proposition 1.

Supplementary Materials

The supplementary materials contain the proofs of Propositions 2–4, some technical lemmas with proofs, and the discussion on the asymptotic Gumbel distribution for the Nadaraya-Watson kernel estimator.

Acknowledgment

The authors are grateful to the Editor, an associate editor and three referees for their valuable and constructive comments which substantially improved an earlier version of the paper. Liu's research was supported by NSFC (Grants No.11322107 and No.11431006), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Shanghai Shuguang Program and 973 Program (2015CB856004). Wang's research was supported by an Australian Research Council Discovery Project.

References

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics* **1**, 1071–1095.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Lecture Notes in Statistics Vol. **110**, Springer, New York.

Cai, Z., Li, Q. and Park, J. Y. (2009). Functional–coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101–113.

Chan, N. and Wang, Q. (2014). Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory* **30**, 1110–1133.

Chen, J., Gao, J. and Li, D. (2012). Estimation in semi-parametric regression with nonstationary regressors. *Bernoulli* **18**, 678–702.

Chen, J., Li, D. and Zhang, L. (2010). Robust estimation in a nonlinear cointegration model. *Journal of Multivariate Analysis* **101**, 706–717.

Engle, R. and Granger, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* **55**, 251–276.

Fan J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* **27**, 715–731.

Gao, J., King, M., Lu, Z. and Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* **25**, 1869–1892.

Gao, J., King, M., Lu, Z. and Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics* **37**, 3893–3928.

Gao, J., Kanaya, S., Li, D. and Tjøstheim, D. (2015). Uniform consistency for nonparametric estimators in null recurrent time series. *Econometric Theory* **31**, 911–952.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**, 121–130.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics, Vol. 60, Springer, Berlin.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston.

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis* **27**, 228–254.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis* **12**, 402–414.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model. *The Annals of Statistics* **35**, 252–299.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2010). Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference* **140**, 3619–3626.

Kasparis, I. and Phillips, P. C. B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics* **168**, 270–284.

Liu, W. and Wu, W. B. (2010). Simultaneous nonparametric inference of time series. *The Annals of Statistics* **38**, 2388–2421.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* **90**, 1257–1270.

Wang, Q. (2014). Martingale limit theorems revisit and non-linear cointegrating regression. *Econometric Theory* **30**, 509–535.

Wang, Q. and Chan, N. (2014). Uniform convergence for a class of martingales with applications in non-linear cointegrating regression. *Bernoulli* **20,** 207–230.

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710–738.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* **77**, 1901–1948.

Wang, Q. and Phillips, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* **27**, 235–259.

Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics* **40**, 727–758.

Wang, Q. and Wang, R. (2013). Non-parametric cointegrating regression with NNH errors.

Econometric Theory **29**, 1–27.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, 69, 391–410.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society B* **60**, 797–811.

Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis* **101**, 1656–1680.

Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics* **36**, 1854–1878.

Department of Mathematics, The University of York, YO10 5DD, York, UK.

E-mail: degui.li@york.ac.uk

Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China.

E-mail: weidongl@sjtu.edu.cn

School of Mathematics and Statistics, The University of Sydney, NSW 2006, Sydney Australia.

E-mail: qiying@maths.usyd.edu.au

Department of Statistics, The University of Chicago, Chicago, Illinois 60637, U.S.A.

E-mail: wbwu@galton.uchicago.edu

(Received June 2015; accepted July 2016)